

Genome-Wide Copy Number Variation Analysis in Extended Families and Unrelated Individuals Characterized for Musical Aptitude and Creativity in Music

Liisa Ukkola-Vuoti^{1*}, Chakravarthi Kanduri^{1,2*}, Jaana Oikonen¹, Gemma Buck³, Christine Blancher³, Pirre Raijas⁴, Kai Karma⁵, Harri Lähdesmäki², Irma Järvelä¹

1 Department of Medical Genetics, University of Helsinki, University of Helsinki, Finland, **2** Department of Information and Computer Sciences Aalto University, Espoo, Finland, **3** Wellcome Trust Centre for Human Genetics, Oxford, United Kingdom, **4** Sibelius Academy, DocMus Department, Helsinki, Finland, **5** Sibelius Academy, Department of Music Education, Helsinki, Finland

Abstract

Music perception and practice represent complex cognitive functions of the human brain. Recently, evidence for the molecular genetic background of music related phenotypes has been obtained. In order to further elucidate the molecular background of musical phenotypes we analyzed genome wide copy number variations (CNVs) in five extended pedigrees and in 172 unrelated subjects characterized for musical aptitude and creative functions in music. Musical aptitude was defined by combination of the scores of three music tests (COMB scores): auditory structuring ability, Seashores test for pitch and for time. Data on creativity in music (herein composing, improvising and/or arranging music) was surveyed using a web-based questionnaire. Several CNVRs containing genes that affect neurodevelopment, learning and memory were detected. A deletion at 5q31.1 covering the protocadherin- α gene cluster (*Pcdha 1-9*) was found co-segregating with low music test scores (COMB) in both sample sets. *Pcdha* is involved in neural migration, differentiation and synaptogenesis. Creativity in music was found to co-segregate with a duplication covering glucose mutarotase gene (*GALM*) at 2p22. *GALM* has influence on serotonin release and membrane trafficking of the human serotonin transporter. Interestingly, genes related to serotonergic systems have been shown to associate not only with psychiatric disorders but also with creativity and music perception. Both, *Pcdha* and *GALM*, are related to the serotonergic systems influencing cognitive and motor functions, important for music perception and practice. Finally, a 1.3 Mb duplication was identified in a subject with low COMB scores in the region previously linked with absolute pitch (AP) at 8q24. No differences in the CNV burden was detected among the high/low music test scores or creative/non-creative groups. In summary, CNVs and genes found in this study are related to cognitive functions. Our result suggests new candidate genes for music perception related traits and supports the previous results from AP study.

Citation: Ukkola-Vuoti L, Kanduri C, Oikonen J, Buck G, Blancher C, et al. (2013) Genome-Wide Copy Number Variation Analysis in Extended Families and Unrelated Individuals Characterized for Musical Aptitude and Creativity in Music. PLoS ONE 8(2): e56356. doi:10.1371/journal.pone.0056356

Editor: Andrew Dewan, Yale School of Public Health, United States of America

Received: October 2, 2012; **Accepted:** January 14, 2013; **Published:** February 27, 2013

Copyright: © 2013 Ukkola-Vuoti et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The work was financially supported by the Academy of Finland (grant 13371) and the Biocentrum Helsinki Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: liisa.ukkola@helsinki.fi

These authors contributed equally to this work.

Introduction

Like intelligence or language, music perception and practice are complex cognitive functions of the human brain. Twin studies have shown evidence for genetic overlap between different cognitive abilities such as learning, reading and mathematics [1]. Neurobiological studies [2,3,4,5], studies on animals [6,7,8], and human infants [9,10,11,12] have provided evidence on biological basis of music perception. There is an abundance of data about the neurophysiological effects of music perception and practice on the human brain [2,3,4,5,12]. However data on the effects at the cellular is so far missing.

Recent studies have shown a substantial genetic component in music perception including absolute pitch [13], congenital amusia

[14], auditory structuring ability [15,16,17] and musical ability [18]. Until now, little evidence for the molecular genetic background of musical aptitude has been obtained. In the absence of such evidence, we and others have performed both genome wide analyses and candidate gene studies in musical traits [13,15,16,17,18,19,20]. Intriguingly, genome-wide analyses performed separately in Finnish and Mongolian populations with different music phenotypes (musical aptitude and musical ability) revealed linkage in the partly overlapping genetic regions at chromosome 4q [15,18]. Previously, such candidate genes as *AVPR1A*, *SLC6A4*, *UNC5C* and *UGT8* have been suggested for musical abilities [15,16,17,18]. These preliminary molecular studies support the hypothesis that musical aptitude is the result

of currently unknown number of genomic variations, environment, and their complex interactions.

Variations in the human genome range from large chromosomal anomalies (size >2–5 Mb) to single nucleotide polymorphisms (SNPs) (size from 1 to 700 bp) [21]. Copy number variations (CNVs) are structural genomic variants arising from deletions or duplications of the genomic region. CNVs show high variability in the human genome and have been suggested to have multiple effects on gene function, evolution and disease risk [22,23]. To date, CNVs have been shown to have an important role in cognitive function of the human brain in neuropsychiatric disorders [23,24,25,26] and in common complex diseases [27,28]. Lately, attention has been paid to CNVs underlying normal human traits like height and intelligence [29,30,31]. Most of these studies have been performed in case-control settings and have focused on diseases [32] whereas family-based studies on normal cognitive traits are rare.

In order to obtain further understanding about the biological basis of music perception, we conducted a genome-wide survey of CNVs in five multigenerational families and in 172 unrelated subjects who were characterized for musical aptitude and creativity in music. The advantages of a family-based setting used here over population-based include better control of population stratification, enrichment of rare variants and the ability to discriminate variants co-segregating with the trait [33,34]. High mutation rate is resulting in a large number of *de novo* variations [23]. The family based approach is useful as it offers validation for CNV each time it is inherited.

Materials and Methods

Ethics statement

The study was approved by The Ethical Committee of Helsinki University Central Hospital and was conducted in accordance with the Declaration of Helsinki. An informed consent was obtained from all subjects.

Study material

The age of the participants varied from 18 to 60 years. To minimize the age-related affects on CNV accumulation and various music tests, we made sure that all the individuals participating in this study were aged ≤ 60 years.

The family pedigrees are shown in Figure 1. The families consist of from 28 to 38 genotyped members whose combined music test scores (COMB) and creative functions in music are known (Table 1). The families participate in the study where genetic background of music perception and practice is studied [15,16,17]. Of the study families, family 6 has the lowest mean COMB scores (116.98), and creative functions in music were reported by 10%. Family 14 has the highest mean COMB scores (128.21) and 46% of the members reported creative functions in music.

Further, a sample set of 172 unrelated subjects were selected to analyze the effect of CNV burden and the enrichment of CNVs in opposite phenotypes of musical aptitude and creative functions in music (Table 1). The unrelated subjects originate from the genetically isolated Finnish population that has experienced multiple bottlenecks in its population history [35] so we cannot fully exclude their distant relatedness. In the questionnaire participants could report if any close relatives participated in the study. The individuals in the unrelated sample set were selected if they did not report any relatives in the study or did not show relatedness in identity by descent (IBD) analysis. No medical information was available from the participants but as far as we

know, they are healthy. The recruitment process has been described in our earlier studies [15,16,17].

Phenotypes

Musical aptitude of each participant was defined using three music tests: the auditory structuring ability test (Karma Music test, KMT) [36] and Carl Seashore's pitch and time discrimination subtests (SP and ST respectively) [37] described previously by Pulli et al. [15] and Ukkola et al. [16]. Shortly, KMT contains 40 items that measure recognition of melodic contour, grouping, relational pitch processing, and gestalt principles, the same potentially innate musical cognitive operations reported by Justus & Hutsler [38]. In contrast, Seashore's tests each contain 50 items that measure simple sensory capacities, such as the ability to detect small differences in tone pitch or duration that are necessary in music perception. A combined music test score (COMB), was computed as the sum of the separate scores of the three individual test results (range from 75 to 150 scores), as described earlier [16,17]. There is moderate correlation between the three music tests scores: 0.61 between KMT and SP, 0.42 between KMT and ST and 0.38 between SP and ST ($P < 0.0001$ for all three) [17]. The SP has shown the highest heritability (52%) whereas the heritability estimates for KMT, ST and COMB are 39%, 10% and 44% respectively [16,17]. The reliabilities of the music tests range from 0.78 to 0.91. [15]. The age had a significant effect on COMB scores ($F = 19.24$, $df = 169$, $p = 2.02 \times 10^{-5}$). The R-square of 0.09 suggests that about 9% of the variability in COMB scores could be explained by age.

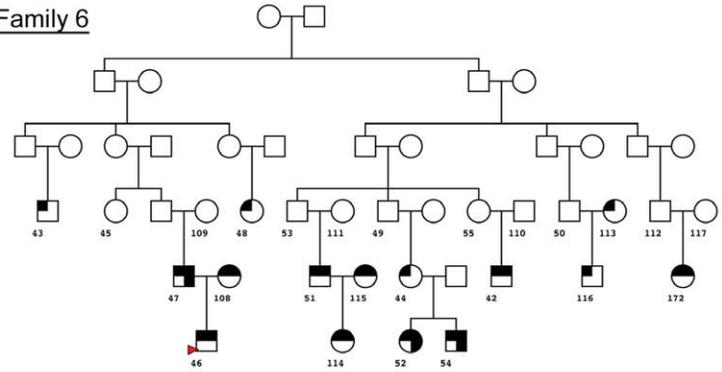
Background information, e.g. about the participant's creativity in music was collected using a web-based self-report questionnaire as described earlier [16,17]. Here the creative functions in music were defined as having one or several of the following: composing, improvising or arranging music. The questionnaire is available from the authors on request.

In this study, the music phenotypes were explored in opposite phenotypes. Technically, we compared the distribution of CNVs among subjects belonging to the two groups of COMB scores. Based on the lower and upper quartiles of COMB scores, the subjects were divided into (1) low and (2) high COMB scoring group within each family and also unrelated sample (Table 2) (3) "Creative phenotype" here means a subject reporting one or several of the aforementioned creative functions in music and (4) "noncreative phenotype" was applied if not reporting any of these activities. Data about creativity was available in family 6 from 48%, family 13 from 63%, family 14 from 75%, family 15 from 53%, family 17 from 74% and unrelated data from 73% of the family members participating in the study (Table 2).

Genotyping

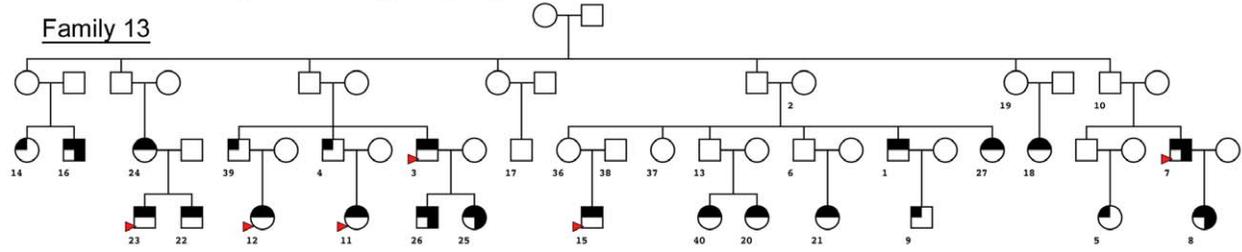
Peripheral blood samples for genomic DNA extraction were drawn from the participants over 12 years of age (no cell lines were used). 200 ng of DNA from each subject was genotyped using the Illumina Infinium Human OmniExpress-12v1 beadchip, which contains an approximate number of 733,202 markers per sample for SNP and CNV analyses. Samples were genotyped with an average overall call rate of 99.54%. Normalized signal intensity data was obtained through Illumina BeadStudio software. Information on Log_2 R ratios, B allele frequencies, markers and chromosomal coordinates from each sample were used for CNV identification. This is the first study using the SNP data obtained from the analysis.

Family 6

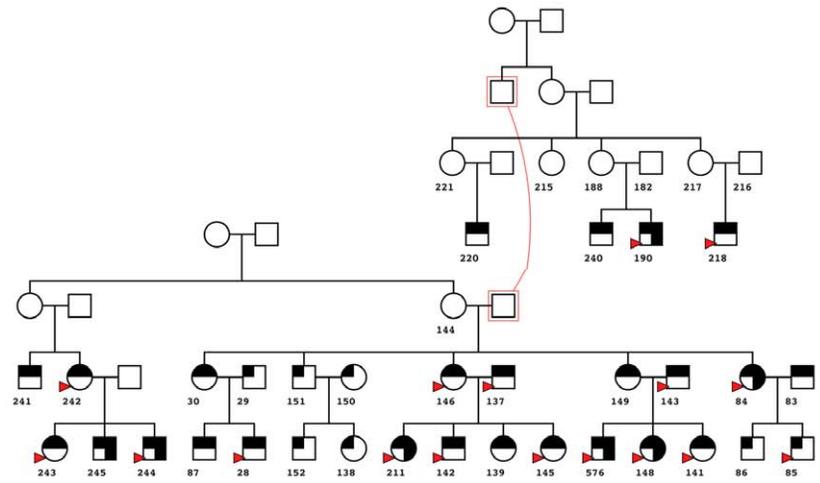


COMB Music test scores:
 ■ LOW
 ◻ MEDIUM
 ◼ HIGH
 ▶ Creative functions in music

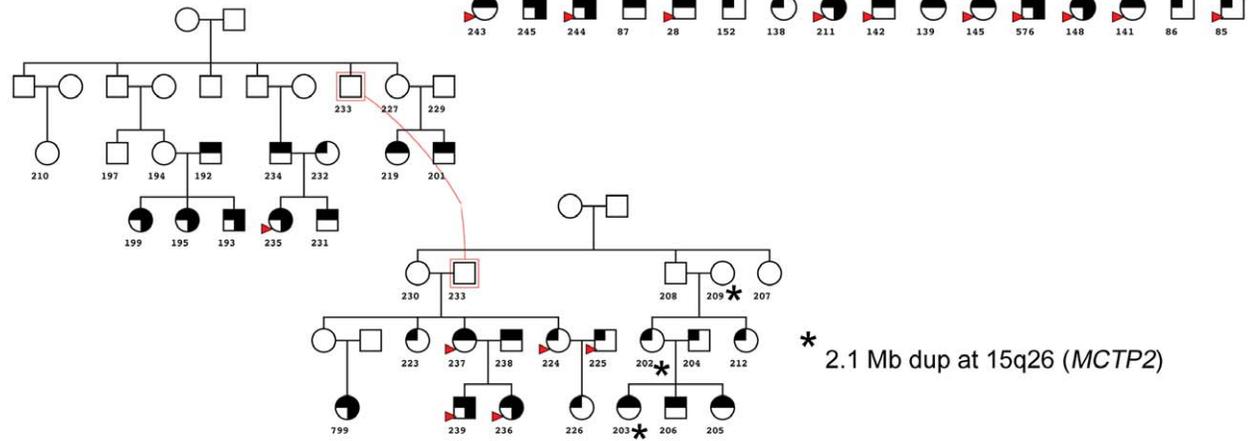
Family 13



Family 14



Family 15



Family 17

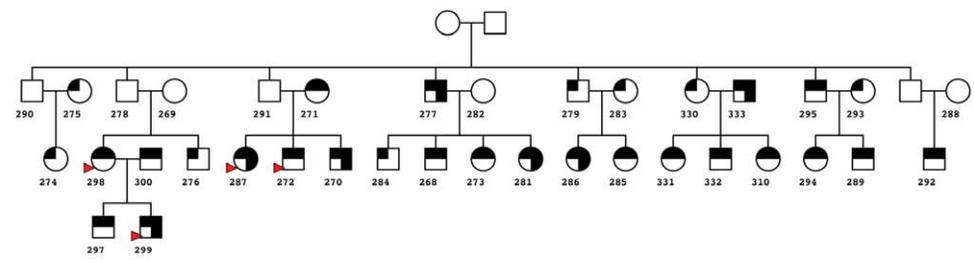


Figure 1. The families of the study. Circles represent females, squares males and genotyped individuals were marked with subject numbers.
doi:10.1371/journal.pone.0056356.g001

CNV identification and quality control

We mapped all the probe coordinates in this study to human genome build GRCh37/hg19. CNVs were identified using two Hidden Markov Model (HMM)-based algorithms: PennCNV and QuantiSNP. At present, both of these algorithms constitute the most reliable set-up for CNV detection using Illumina platforms [39]. Additionally, this multi-algorithm approach increases the confidence of CNV calls and reduces false positives [39,40,41]. For familial data, using the trio-based CNV calling in PennCNV, we validated the inheritance status of CNVs. Further, to ensure good quality of the data, we followed stringent quality control criteria at both sample-level and CNV-call level. Specifically, we categorized certain samples as outliers and eliminated them if they meet any of the following criteria: (a) Call rate below the call threshold of 98%. (b) Standard deviation of Log_2 R ratio more than 0.15 (c) Standard deviation of B allele frequency more than 0.05 (d) Waviness factor outside the limits of -0.04 and 0.04 (4) BAF drift more than 0.002 (e) Number of detected CNVs per sample more than 50. At the level of CNV-calls, CNVs identified using both PennCNV and QuantiSNP were retained only if their size was larger than 10 bp and their log Bayesian factor was more than 10 respectively. A CNV detected by both algorithms was merged into a single consistent-call (consensus) by using the outermost boundaries defined by either of the algorithm, irrespective of their size of overlap. We retained only such consistent calls for further analyses.

Statistical analyses

To unravel the impact of CNVs on musical aptitude and creative functions in music, we analyzed the detected CNVs using three different approaches: First, we analyzed the inheritance of CNVs in extended pedigrees and their penetrance in contrasting phenotypes, i.e. high COMB vs low COMB and creative vs non-creative subjects. The main objective of this analysis is to find the highly penetrant CNVs that are private to a specific phenotype. For this, CNVRs larger than 10 kb were ranked in each family in individuals of contrasting phenotypes (Table 2). This straightforward ranking method ranks a CNV based on its frequency count in the “affected” individuals of each family. For example, the CNV that appears the most frequent in musically creative individuals of a family is given a higher rank for musical creativity within that family. This ranking method was described earlier in Karlsson et al. [26]. Second, we investigated the effect of CNV burden on 172 unrelated individuals characterized for the aforementioned phenotypes. For this, we tested if there was an increased burden of CNVs in opposite phenotypes concerning

number of CNVs and CNV size using a two-sided Fisher’s exact test. Third, we checked if any particular CNV was present with increased frequency in the opposite phenotypes of unrelated individuals using a two-sided Fisher’s exact test; p -value <0.05 . All these Statistical analyses were performed using R (statistical computing environment, <http://www.r-project.org/>), SPSS predictive analytics software version 20, PLINK and custom scripts.

Results

CNVRs in musical aptitude (COMB scores)

In the five study families all CNVRs were ranked for COMB scores and are shown in Supplementary material (Table S1). High ranking CNVRs shared by individuals in at least two families for the COMB scores are shown in Table 3. One CNVR was ranked high in high COMB scores. A total of 67% of family members with high COMB scores in families 6 and 14 members carried a deletion at 1q21.2 (Table 3). The region contains genes *FCGR1C* and *LOC388692* and has previously been linked with neurodevelopmental disorders (schizophrenia, autism, ADHD, mental retardation, learning disabilities and dyslexia) [42,43]. A family-specific deletion at 8p23 (*DLGAP2*), previously related to autism [39], was also ranked high in family 14.

In low COMB scoring individuals one CNVR was ranked high in two different families. This was a deletion at 5q31.3 found in 54% of low COMB scoring individuals in families 14 and 15. The region contains the protocadherin- α gene cluster 1–9 (*Pcdha 1-9*) consisting of 14 tandemly arranged genes [44]. Protocadherins are composed of α -, β - and γ -clusters arranged in tandem on human chromosome 5 and expressed in vertebrate brain [45,46]. The *Pcdha* encode diverse proteins whose functions are involved in axonal projection and in learning and memory. Several other family-specific interesting CNVRs were ranked high for low COMB scores. In family 14 there was a duplication at 17q21.31 that contains the *KANS1* (also called *KIAA1267*) gene. Also, a deletion at 9p21.1 (*LINGO1*) in family 15 was ranked high for low COMB scores.

A number of large family-specific CNVs were detected in low COMB scoring individuals. A novel 2.1 Mb duplication (not found in the Database of Genomic Variants) at 15q26 was inherited in family 15 in three generations (Figure 1, ID 209, 202, 203). Interestingly, this region contains the *MCTP2* gene previously associated with schizophrenia in samples consisting of Norwegian, Swedish and Danish subjects [47]. In family 6 a novel large 1.3 Mb duplication was identified at 8q24.22 (rs4518624–

Table 1. Characteristics of each family and the unrelated sample set (N-value).

Pedigree no.		6	13	14	15	17	unrelated
DNA available		25	32	39	36	38	172
sex	male	12 (48%)	16 (50%)	20 (53%)	15 (42%)	18 (47%)	71 (41%)
	female	13 (52%)	16 (50%)	18 (47%)	21 (58%)	20 (53%)	101 (59%)
COMB music test scores	mean	116.98	120.66	128.21	117.17	118.3	122.41
	range	93.5–144.25	88.25–141.5	82–147	83–147	90–144.8	89–148
Creativity in music	No	18 (90%)	19 (70%)	19 (54%)	22 (79%)	30 (88%)	125 (86%)
	Yes	2 (10%)	8 (30%)	16 (46%)	6 (21%)	4 (12%)	21 (14%)

doi:10.1371/journal.pone.0056356.t001

Table 2. Descriptive statistics for the analyzed music phenotypes in the study material.

	COMB scores		Creative phenotype	
	High (% of total N)	Low (% of total N)	Yes (% of total N)	No (% of total N)
Fam no. 6	>144.2	<115.0		
N	5 (20%)	4 (16%)	1 (4%)	11 (44%)
Fam no. 13	>138.8	<114.5		
N	5 (16%)	5 (16%)	6 (19%)	14 (44%)
Fam no. 14	>141.8	128.5		
N	7 (18%)	7 (18%)	17 (44%)	12 (31%)
Fam no. 15	>137.0	<105.0		
N	6 (17%)	6 (17%)	5 (14%)	14 (39%)
Fam no. 17	>133.2	<111.5		
N	8 (21%)	8 (21%)	4 (11%)	24 (63%)
Unrelated	>138.2	<116.2		
N	40 (23%)	28 (16%)	18 (10%)	108 (63%)

doi:10.1371/journal.pone.0056356.t002

rs9297816) in a subject with low COMB scores (Figure 1, family 6 ID 48). This region contains the genes *ADCY3*, *ASAP1*, *FAM49B* and *GSDMC*. The duplication is located inside the region that has previously been linked with absolute pitch (rs755520–rs2102861) [13] (Figure 2).

No significant excess of large CNVs or CNV burden were detected in unrelated individuals with high or low COMB scores (Table S2).

CNVRs in the creative phenotype

In the five study families CNVRs were ranked for creative phenotype and are shown in the Supplementary material (Table S3). A total of five CNVRs were ranked high in the creative phenotype in at least two different families (Table 3). A deletion at 5p15.33 containing the gene *ZCHHC11* was present in 48% and a duplication at 2p22.1 containing the gene *GALM* was present in 27% of creative family members (family 14 and 17; family 14 and 15, respectively). *GALM* encodes galactose mutarotase which is functioning in serotonin metabolism [48,49]. Curiously, we and others have studied the association of the serotonin transporter (5-HTT) in music related phenotypes [16,17,19,20,50]. A duplication

at 10q11.22 was found in 24% and a deletion at 3p22.2 in 20% of the creative phenotype subjects in two families each (families 14 and 17; families 13 and 17, respectively). 19% of the members in families 14 and 17 carried an inherited deletion at 5p15.31. Family-specific deletions at 6q12 [51] and at 8p23.3 containing the *DLGAP2* gene, a candidate gene for autism [39], were ranked high in creativity (Table S3). The deletion at 8p23.3 was also ranked high among high COMB scores (Table S3).

Three CNVRs were ranked high for noncreative individuals from at least two different families. A deletion at 2p12 was carried by 31% and a deletion at 3p14.1 by 21% of the noncreative subjects in two families (family 14 and 17; family 13 and 17, respectively). 19% of non-creative individuals from three families (6, 14 and 17) carried a deletion at 3q28 containing the *CCDC50* gene.

In the unrelated sample set no significant excess of large CNVs or CNV burden were detected in the creative or non-creative phenotypes (Table S2).

Table 3. High-ranked CNVRs for COMB scores and creative phenotype in the analyzed families.

	freq. in the phenotype	Chr region	Chr: start-end	Event type	Genes	Families
High COMB music test scores	67%	1q21.2	1:149039031–149388389	Loss	<i>FCGR1C</i> , <i>LOC388692</i>	6, 14
Low COMB music test scores	54%	5q31.3	5:140225908–140237548	Loss	<i>Protocadherin alpha gene cluster</i>	14, 15
Creative phenotype	48%	5p15.33	5:788646–840717	Loss	<i>ZDHHC11</i>	14, 17
	27%	2p22.1	2:38955129–38977612	Gain	<i>GALM</i>	14, 15
	24%	10q11.22	10:47412588–47703869	Gain	<i>ANTXR1</i>	14, 17
	20%	3p22.2	3:37979882–37986249	Loss	<i>CTDSPL</i>	13,17
	19%	5p15.31	5:8258881–8260630	Loss	-	14,17
Noncreative phenotype	31%	2p12	2:76941049–76949101	Loss	-	14,17
	21%	3p14.1	3:65191847–65214685	Loss	-	13,17
	19%	3q28	3:191065392–191072060	Loss	<i>CCDC50</i>	6, 14, 17

doi:10.1371/journal.pone.0056356.t003

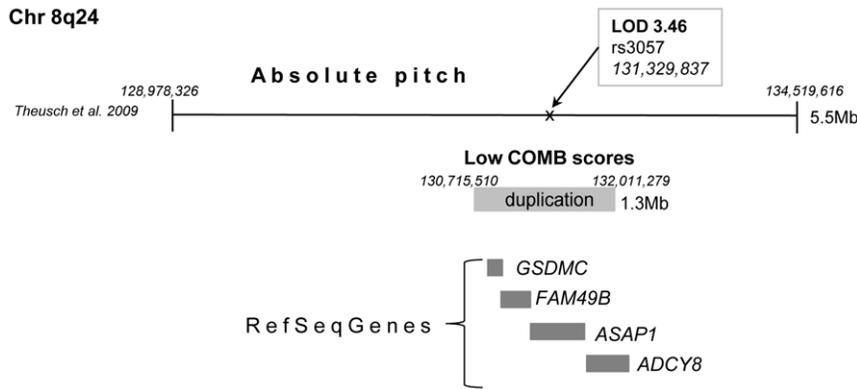


Figure 2. 1.3 Mb duplication in the region previously linked with absolute pitch.
doi:10.1371/journal.pone.0056356.g002

Association of CNVRs with music phenotypes in the unrelated sample set

After family-based ranking of CNVRs in COMB scores and creative phenotype was completed, the association analysis in the unrelated sample set was performed (Table 4 and Table 5). A duplication at 12p11.21 was slightly enriched in individuals with high COMB scores (Fisher $p = 0.0385$), while a deletion at 3p14.1 was observed in individuals with low COMB scores (Fisher $p = 0.0322$). No known genes were located in these CNVRs. The deletion at 3p14.1 was also found in the noncreative family members in CNVR ranking analysis (Table 3). A deletion at 5q31.1 containing *Pcdh* was seen in 7% of low COMB and in none of the high COMB subjects.

Seven different CNVRs showed very suggestive association with the creative phenotype and one CNVR to the non-creative phenotype (Table 5). Of the mildly associated CNVRs, seven were deletions and one was duplication. Interestingly, the deletion at 5p15.33, for creative subjects was found in both the family-based (48%) and unrelated (22%) sample sets. The most promising, though preliminary, associations were seen between the creative phenotype and deletions at 6q14, 7q11 and 7q33 (Fisher $p = 0.0194$). 11% of creative vs. 0% of noncreative subjects were carrying each of these deletions. In previous studies, 6q14, containing the genes *MYO6* and *SEN6*, has been associated with intellectual disability and language delay [51]. The region 7q11, containing genes *INTS4L1* and *ZNF92*, is a candidate locus for epilepsy [52]. Interestingly, regions near to 7q33, containing the gene *EXOC4*, has previously been linked to speech and language disorders [53] and dyslexia [54]. Further, a deletion at 12p11 (17% creative vs. 2% non-creative), and a 70 kb deletion at 6p21 (28% creative vs. 8% non-creative) were weakly associated with creative functions in music ($p = 0.0207$ and $p = 0.0298$, respectively). A duplication at 7p12 was enriched in 22% of creative individuals. For the non-creative group, a deletion at 3p12, locating near to a susceptibility gene for reading and language

disorder *ROBO1* [55], was more common compared to the creative group (24% non-creative vs. 0% creative).

We acknowledge that because of the multiple testing problem the associations detected here are only suggestive and preliminary.

Discussion

Low music test scores

In the genome wide analysis of CNVs in musical traits, several CNVRs containing genes that affect neurodevelopment, learning and memory were detected. The most relevant CNVR found here was a deletion at 5q31.1 covering the protocadherin- α gene cluster 1–9 (*Pcdha 1-9*). This deletion was found from both sample sets in subjects with low music test scores (COMB): 54% of the cases from two families (Table 3) and 7% from the unrelated subjects. There is an abundance of data about the neurophysiological effects of music perception and practice on the human brain [2,3,4,5,12]. However data on the effects at the cellular levels is so far missing. *Pcdha* is involved in neural migration, differentiation and synaptogenesis [56,57,58]. Katori and colleagues [60] discovered that *Pcdhas* are important in maturation of serotonergic projections in all or most of the brain regions. Previously, Fukuda et al. [59] showed abnormalities in *Pcdha* mutant mice in abilities important for musical aptitude, that is learning and memory. The human serotonin transporter gene (*SLC6A4*) together with arginine vasopressin receptor gene (*AVPR1A*) polymorphisms has been reported to associate with artistic creativity in professional dancers [50] and with short-term musical memory [19]. Based on the normal function of the brain, *Pcdha* may be a relevant candidate gene affecting music perception and practice. In our previous studies, *AVPR1A*, also related to learning and memory, was associated with musical aptitude and listening to music [16,17].

Absolute pitch (AP), which is a rare ability to recognize the pitch of a musical tone without a reference pitch, was previously linked on chromosome 8q24.21 [13]. In our study, a 1.3 Mb long

Table 4. Association of CNVRs with music test scores in the unrelated dataset.

	High COMB (N = 40)	Low COMB (N = 28)	Chr region	Chr: start-end	Event type	Genes	p-value
COMB music test scores	8 (20%)	13 (46%)	3p14.1	3: 65191847–65214685	loss	-	0.0322
	6 (15%)	0 (0%)	12p11.21	12: 31266287–31409778	gain	-	0.0385

doi:10.1371/journal.pone.0056356.t004

Table 5. Association of CNVRs with creative functions in music in the unrelated dataset.

	Creative (N = 18)	Non-creative (N = 108)	Chr region	Chr: start-end	Event type	Genes	p-value
Creative phenotype	2 (11%)	0 (0%)	6q14.1	6: 76218758–76552776	loss	<i>MYO6.SENP6</i>	0.0194
	2 (11%)	0 (0%)	7q33	7: 133146663–133235334	loss	<i>EXOC4</i>	0.0194
	2 (11%)	0 (0%)	7q11.21	7: 64679561–65326821	loss	<i>INTS4L1.ZNF92</i>	0.0194
	3 (17%)	2 (2%)	12p11.21	12: 32004170–32056577	loss	-	0.0207
	0 (0%)	26 (24%)	3p12.3	3: 75419736–75655870	loss	<i>FAM86DP</i>	0.0232
	4 (22%)	5 (5%)	7p12.1	7: 52733291–52743803	gain	-	0.0237
	5 (28%)	9 (8%)	6p21.32	6: 32493790–32560859	loss	-	0.0298
	4 (22%)	6 (6%)	5p15.33	5: 674921–840717	loss	<i>ZDHHC11</i>	0.0359

doi:10.1371/journal.pone.0056356.t005

duplication was located in the core linkage region of AP (Figure 2) in a subject with low COMB scores. The overlapping region contains gene *ADCY8* (*adenylate cyclase 8*) that is associated with synaptic plasticity, short-term memory performance [60] and with bipolar disorder [61]. Additionally, genes *GSDMC*, *FAM49B* and *ASAP1* lie in the region of duplication (RefSeq). Disorders caused by CNVs of dosage sensitive genes critical for the physiological function of nervous system have been identified [62]. Further, large duplications have shown harmful effects on neurodevelopment [43,63]. Music test scores used in our study tests e.g. subject's ability to detect differences in pitch. Although there are no studies about the overlap of AP and the phenotypes studied here, it is intriguing to speculate whether the duplication in the AP region had a harmful effect on pitch perception accuracy in the multifaceted phenotype of musical aptitude.

Creativity and psychiatric disorders

Creativity is an ability to produce work that is not only original but appropriate for the situation in which it occurs [64]. In our previous study creativity in music (here composing, improvising or arranging) showed substantial heritability [16]. Here, a duplication at 2p22.1, containing the *GALM* gene, was co-segregating with creative phenotype. *GALM* encodes galactose mutarotase which has been reported to increase serotonin release and membrane trafficking of the human serotonin transporter (5-HTT) [48]. *GALM* has been associated with serotonin transporter binding potential in the human thalamus [48], the region that is important for the music perception process [65]. Serotonin metabolism is disturbed in mood disorders e.g. depression [49]. Recently, serotonin transporter gene (*SLC6A4*) has been associated not only with psychiatric disorders but also with musical aptitude [16], attending choral singing [19] and creative dance performance [50]. It is known that genetic polymorphisms that are related to psychiatric disorders may have a positive impact on cognitive traits like creativity, IQ, and working memory [50,66,67,68]. The other side of the coin is that highly creative individuals may have an elevated risk for certain types of psychiatric disorders. Kyaga et al. [68] suggested co-segregation of mental disorders with creativity in families. The risk alleles may act like plasticity genes, resulting in that carriers are more responsive to both positive and negative environmental experiences than others [69].

Music is non-verbal communication that is able to evoke emotions which are unique in intensity and state [70,71]. Listening to and/or playing music is environmental stimuli that has multiple measurable effects on brain structure and function. Neurophysi-

ological and brain imaging studies have discovered that music induces synaptic plasticity, e.g. active training and practicing music has been shown to enlarge cortical presentations in the somatosensory and auditory domains in professional musicians [72]. However, little is known about the effect of music on the brain function at the cellular level. In our study, a large 2.1 Mb long duplication in the region of the *MCTP2* gene, previously reported in schizophrenia [47] was co-segregated in three generations with low or average music test scores. *MCTP2* is involved in cellular signal transduction and synaptic functions by Ca²⁺ binding [73]. In schizophrenia, Ca²⁺ binding is altered in the prefrontal cortex [74]. This same brain region is also important for recognizing emotions in music [75], which makes us hypothesize the effect of this large duplication to the function of prefrontal cortex in music perception [41,62,63].

In our study, a deletion at 3p14.1 was ranked high in both family and unrelated data for different phenotypes; in the family data the non-creative phenotype, and in the unrelated sample set low music test scores. In our previous study high music test scores were associated with creativity in music ($p < 0.0001$) [16]. This may support the finding that deletion at 3p14.1 is associated with both low music test scores and the non-creative phenotype.

The size of the CNVRs

Currently, the standardized practices and gold standard for CNV studies, especially across different size spectrum, are lacking. Although several studies raised the issue of potential false discoveries among smaller CNVs (<100 kb), it remains intriguing for the researchers to study smaller CNVs because of their well-acknowledged functional impact [76,77]. Moreover, CNVs >100 kb were previously suggested to have deleterious effects [23] and in general they have been heavily implicated in neurodevelopmental disorders. As this study focuses on a normal trait, our presumption allowed us to consider smaller CNVs also. However, we tried to minimize the false discoveries by using a multi-algorithm approach which has been promising in previous studies [39,40,41].

Concluding Remarks

Here we report the results of the first genome wide CNV survey for music related phenotypes; musical aptitude and creative functions in music using both a family-based approach and case-control study. The advantages of this study are the use of both family-based and sporadic data. Moreover, all samples were genotyped at the same time using the same platform, increasing

the reliability of the analyses. However, there are several limitations in our study. Definition of the phenotype here covers only a small portion of the multifaceted phenotype of music perception and practice. Being aware of the quantitative nature of musical aptitude, division of the phenotype to high COMB and low COMB groups is somewhat artificial. Also, the sample size is relatively small and the participants have not been screened for neurocognitive deficits. Consequently, the identified CNVs cannot be excluded here as being potentially predisposing for neuropsychiatric conditions. Because of the small sample size only suggestive associations were detected. The result, although interesting, is preliminary and replication with a larger sample set is needed.

The perception of sounds begins in the cochlea, the auditory portion of the inner ear, but the actual perceiving, processing and creating music takes place at multiple sites and elicits different functions of the brain [4,5,70,78,79,80,81]. The thalamus is responsible for projection of the sound information into the auditory cortex where more specific information about acoustic signal, like pitch height, chroma, intensity and timbre, is further extracted [4]. CNVs found in this study contain several genes that are expressed in brain regions where music is perceived, including hippocampus, thalamus and prefrontal cortex. Further studies are needed to survey the mechanisms of the detected CNVs and genes for the human brain and further to music perception and practice.

References

- Haworth CM, Kovas Y, Harlaar N, Hayiou-Thomas ME, Petrill SA, et al. (2009) Generalist genes and learning disabilities: a multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. *J Child Psychol Psychiatry* 50(10):1318–25.
- Elbert T, Pantev C, Wienbruch C, Rockstroh B, Taub E (1995) Increased cortical representations of the fingers of the left hand in string players. *Science* 270: 305–05.
- Blood A, Zatorre R (2001) Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *PNAS* 98: 11818–11823.
- Koelsch S (2005) Neural substrates of processing syntax and semantics in music. *Curr Opin Neurobiol* 15(2):207–12.
- Salimpoor VN, Benovoy M, Larcher K, Dagher A, Zatorre RJ (2011) Anatomically distinct dopamine release during anticipation and experience of peak emotion to music. *Nat Neurosci* 14(2):257–62.
- Porter D, Neuringer A (1984) Music discriminations by pigeons. *J Exp Psych. Animal behavior processes* 10(2):138–148.
- Watanabe S, Sato K (1999) Discriminative stimulus properties of music in Java sparrows. *Behav Process* 47:53–57.
- McDermott J, Hauser MD (2007) Nonhuman primates prefer slow tempos but dislike music overall. *Cognition* 104(3):654–668.
- Zentner MR, Kagan J (1996) Perception of music by infants. *Nature* 383(6595):29.
- Trehub SE (2001) Musical predispositions in infancy. *Ann N Y Acad Sci* 930:1–16.
- Hauser MD, McDermott J (2003) The evolution of the music faculty: a comparative perspective. *Nat Neurosci* 6:663–668.
- Perani D, Saccuman MC, Scifo P, Spada D, Andreolli G, et al. (2010) Functional specializations for music processing in the human newborn brain. *Proc Natl Acad Sci U S A* 107(10):4758–63.
- Theusch E, Basu A, Gitschier J (2009) Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity. *Am J Hum Gen* 85(1):112–119.
- Peretz I, Cummings S, Dube MP (2007) The genetics of congenital amusia (tone deafness): a family-aggregation study. *Am J Hum Genet* 81:582–588.
- Pulli K, Karma K, Norio R, Sistonen P, Goring HH, et al. (2008) Genome-wide linkage scan for loci of musical aptitude in Finnish families: evidence for a major locus at 4q22. *J Med Genet* 45(7):451–456.
- Ukkola LT, Onkamo P, Raijas P, Karma K, Järvelä I (2009) Musical Aptitude Is Associated with AVPR1A-Haplotypes. *PLoS ONE* 4(5): e5534. doi:10.1371/journal.pone.0005534
- Ukkola-Vuoti L, Oikkonen J, Onkamo P, Karma K, Raijas P, et al. (2011) Association of the arginine vasopressin receptor 1A (AVPR1A) haplotypes with listening to music. *J Hum Genet* 56(4):324–9.
- Park H, Lee S, Kim H-J, Ju Y S, Shin J-Y, et al. (2012) Comprehensive genomic analyses associate UGT8 variants with musical ability in Mongolian population. *J Med Genet* 49(12): 747–752.
- Granut R, Frankel Y, Gritsenko V, Lerer E, Gritsenko I, et al. (2007) Provisional evidence that the arginine vasopressin 1a receptor gene is associated with musical memory. *Evolution and Human Behavior* 28:313–318.
- Morley AP, Narayanan M, Mines R, Molokhia A, Baxter S, et al. (2012) AVPR1A and SLC6A4 Polymorphisms in Choral Singers and Non-Musicians: A Gene Association Study. *PLoS ONE* 7(2): e31763. doi:10.1371/journal.pone.0031763
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, et al. (2006) Global variation in copy number in the human genome. *Nature* 444(7118):444–454.
- Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315:848–853.
- Itsara A, Cooper GM, Baker C, Girirajan S, Li J, et al. (2009) Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 84 (2):148–161.
- Magri C, Sacchetti E, Traversa M, Valsecchi P, Gardella R, et al. (2010) New copy number variations in schizophrenia. *PLoS One* 5(10):e13422.
- Salyakina D, Cukier HN, Lee JM, Sacharow S, Nations LD, et al. (2011) Copy number variants in extended autism spectrum disorder families reveal candidates potentially involved in autism risk. *Plos One* 6:e26049.
- Karlsson R, Graae L, Lekman M, Wang D, Favis R, et al. (2012) MAG11 copy number variation in bipolar affective disorder and schizophrenia. *Biol Psych* 71(10):922–30. Epub 2012 Feb 28.
- Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatos E, et al. (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464(7289):713–20.
- Park JH, Lee S, Yu HG, Kim JL, Seo JS (2012) Copy number variation of age-related macular degeneration relevant genes in the Korean population. *PLoS One*. 2012; 7(2): e31243.
- Li X, Tan L, Liu X, Lei S, Yang T, et al. (2010) A genome wide association study between copy number variation (CNV) and human height in Chinese population. *J Genet Genomics* 37:779–785.
- Dauber A, Yu Y, Turchin MC, Chiang CW, Meng YA, et al. (2011) Genome wide association of copy number variation reveals an association between short stature and the presence of low-frequency genomic deletions. *Am J Hum Genet* 89:751–759.
- Yeo RA, Gangestad SW, Liu J, Calhoun VD, Hutchinson KE (2011) Rare copy number deletions predict individual variation in intelligence. *PLoS ONE* 6:e16339.
- Girirajan S, Campbell CD, Eichler EE (2011) Human copy number variation and complex genetic disease. *Annu Rev Genet* 45:203–26.
- Norio R (2003) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. *Hum Genet* 112:457–69.
- Antoniou AC, Easton DF (2003) Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet Epidemiol* 25(3): 190–202.
- Knight S, Uh HW, Martinez M (2009) Summary of contributions to GAW Group 15: family-based samples are useful in identifying common polymorphisms associated with complex traits. *Genet Epidemiol* 33 Suppl 1: S99–104.

Supporting Information

Table S1 All CNVRs ranked for COMB scores in the five study families.

(XLS)

Table S2 Rare CNV burden analysis by event type and size in High COMB VS low COMB subjects.

(DOCX)

Table S3 All CNVRs ranked for creative functions in music in the five study families.

(XLS)

Acknowledgments

We thank the participants for their collaboration. Minna Varhala, RN, and Minna Ahvenainen, MSc, are thanked for excellent technical help and language revision.

Author Contributions

Conceived and designed the experiments: CK LUV HL IJ. Performed the experiments: GB CB. Analyzed the data: CK LUV. Contributed reagents/materials/analysis tools: PR JO LUV KK. Wrote the paper: LUV CK IJ.

36. Karma K (2007) Musical Aptitude Definition and Measure Validation: Ecological Validity can Endanger the Construct Validity of Musical Aptitude Tests. *Psychomusicology* 19(2): 79–90.
37. Seashore CE, Lewis D, Saetveit JG (1960) Seashore measures of musical talents. Manual. New York: The Psychological Corp.
38. Justus T, Hustler JJ (2005) Fundamental issues in the evolutionary psychology of music: Assessing innateness and domain specificity. *Music Perception*, 23, 1–27.
39. Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, et al. (2010) Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. *Nucleic Acids Res* 38 (9):e105.
40. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, et al. (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol* 29 (6):512–520.
41. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466 (7304):368–372.
42. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, et al. (2008) Large recurrent microdeletions associated with schizophrenia. *Nature* 455(7210): 232–236.
43. Grayton HM, Fernandes C, Rujescu D, Collier DA (2012) Copy number variations in neurodevelopmental disorders. *Prog Neurobiol* 99(1): 81–91.
44. Noguchi Y, Hirabayashi T, Katori S, Kawamura Y, Sanbo M, et al. (2009) Total expression and dual gene-regulatory mechanisms maintained in deletions and duplications of the Pcdha cluster. *J Biol Chem* 284(46): 32002–32014.
45. Fukuda E, Hamada S, Hasegawa S, Katori S, Sanbo M, et al. (2008) Down-regulation of protocadherin-alpha A isoforms in mice changes contextual fear conditioning and spatial working memory. *Eur J Neurosci* 28(7): 1362–1376.
46. Katori S, Hamada S, Noguchi Y, Fukuda E, Yamamoto T, et al. (2009) Protocadherin- α family is required for serotonergic projections to appropriately innervate target brain areas. *J Neurosci* 29:9137–9147.
47. Zollino M, Orteschi D, Murdolo M, Lattante S, Battaglia D, et al. (2012) Mutations in KANSL1 cause the 17q21.31 microdeletion syndrome phenotype. *Nat Genet* 44(6): 636–638.
48. Djurovic S, Le Hellard S, Kähler AK, Jönsson EG, Agartz I, et al. (2009) Association of MCTP2 gene variants with schizophrenia in three independent samples of Scandinavian origin (SCOPE). *Psychiatry Res* 168(3):256–258.
49. Liu X, Cannon DM, Akula N, Moya PR, Knudsen GM, et al. (2011) A non-synonymous polymorphism in galactose mutarotase (GALM) is associated with serotonin transporter binding potential in the human thalamus: Results of a genome-wide association study. *Mol Psychiatry* 16:584–585.
50. Drevets WC, Thase ME, Moses-Kolko EL, Price J, Frank E, et al. (2007) Serotonin-1A receptor imaging in recurrent depression: replication and literature review. *Nucl Med Biol* 34(7): 865–877.
51. Bachner-Melman R, Dina C, Zohar AH, Constantini N, Lerer E, et al. (2005) AVPR1a and SLC6A4 Gene Polymorphisms Are Associated with Creative Dance Performance. *PLoS Genetics* 1(3): e42.
52. Kousi M, Anttila V, Schulz A, Calafato S, Jakkula E, et al. (2012) Novel mutations consolidate KCTD7 as a progressive myoclonus epilepsy gene. *J Med Genet* 49(6): 391–9.
53. Petrin AL, Giacheti CM, Maximino LP, Abramides DV, Zanchetta S, et al. (2010) Identification of a microdeletion at the 7q33-q35 disrupting the CNTNAP2 gene in a Brazilian stuttering case. *Am J Med Genet A* 152A(12): 3164–3172.
54. Matsson H, Tamminen K, Zuchelli M, Anthoni H, Onkamo P, et al. (2011) SNP variations in the 7q33 region containing DGKI are associated with dyslexia in the Finnish and German populations. *Behav Genet* 41(1): 134–140.
55. Hannula-Jouppi K, Kaminen-Ahola N, Taipale M, Eklund R, Nopola-Hemmi J, et al. (2005) The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet* 1(4): e50.
56. Hasegawa S, Hamada S, Kumode Y, Esumi S, Katori S, et al. (2008) The protocadherin-alpha family is involved in axonal coalescence of olfactory sensory neurons into glomeruli of the olfactory bulb in mouse. *Mol Cell Neurosci* 38(1): 66–79.
57. Pedrosa E, Stefanescu R, Margolis B, Petruolo O, Lo Y, et al. (2008) Analysis of protocadherin alpha gene enhancer polymorphism in bipolar disorder and schizophrenia. *Schizophr Res* 102: (1–3)210–219.
58. Katori S, Hamada S, Noguchi Y, Fukuda E, Yamamoto H, et al. (2009) Protocadherin-alpha family is required for serotonergic projections to appropriately innervate target brain areas. *J Neurosci* 29(29): 9139–9147.
59. Fukuda E, Hamada S, Hasegawa S, Katori S, Sanbo M, et al. (2008) Down-regulation of protocadherin-alpha A isoforms in mice changes contextual fear conditioning and spatial working memory. *Eur J Neurosci* 28(7): 1362–1376.
60. Wong ST, Athos J, Figueroa XA, Pineda VV, Schaefer ML, et al. (1999) Calcium-stimulated adenylyl cyclase activity is critical for hippocampus-dependent long-term memory and late phase LTP. *Neuron* 23:787–798.
61. Zhang P, Xiang N, Chen Y, Sliwerska E, McInnis MG, et al. (2010) Family-based association analysis to finemap bipolar linkage peak on chromosome 8q24 using 2,500 genotyped SNPs and 15,000 imputed SNPs. *Bipolar disord* 12(8): 786–792.
62. Gu W, Lupski JR (2008) CNV and nervous system diseases –what’s new? *Cytogenet Genome Res* 123(1–4): 54–64.
63. Almal SH, Padh H (2012) Implications of gene copy-number variation in health and diseases. *J Hum Genet* 57(1): 6.13.
64. Sternberg RJ, Lubart TI (2006) The concept of creativity: Prospects and paradigms. In: Sternberg RJ, Lubart TI, eds. *Handbook of creativity*. New York: Cambridge University Press. pp 3–15.
65. Koelsch S, Siebel WA (2005) Towards a neural basis of music perception. *Trends Cogn Sci* 9(12): 578–584.
66. Reuter M, Roth S, Holve K, Hennig J (2006) Identification of first candidate genes for creativity: a pilot study
67. Keri (2009) A Promoter Polymorphism of the Neuregulin 1 Gene Is Related to Creativity in People With High Intellectual Achievement. *Psychological Science* 20:1070–1073.
68. Kyaga S, Lichtenstein P, Boman M, Hultman C, Långström N, et al. (2011). Creativity and mental disorder: family study of 300 000 people with severe mental disorder. *BJPsych* 199: 373–379.
69. Belsky J, Jonassaint C, Pluess M, Stanton M, Brummett B, et al. (2009) Vulnerability genes or plasticity genes? *Molecular Psychiatry* 14: 746–754.
70. Menon V, Levitin DJ (2005) The rewards of music listening: response and physiological connectivity of the mesolimbic system. *Neuroimage* 15(1): 175.184.
71. Zentner M., Grandjean D, Scherer KR (2008) Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8: 494–521.
72. Abbott A (2002) Music, maestro, please! *Nature* 416(6876): 12–14.
73. Shin OH, Han W, Wang Y, Sudhof TC (2005) Evolutionarily conserved multiple C2 domain proteins with two transmembrane regions (MCTPs) and unusual Ca²⁺ binding properties. *J Biol Chem* 280(2): 1641–1651.
74. Sakai T, Oshima A, Nozaki Y, Ida I, Haga C, et al. (2008) Changes in density of calcium-binding-protein-immunoreactive GABAergic neurons in prefrontal cortex in schizophrenia and bipolar disorder. *Neuropathology* 28(2) 143–150.
75. Moghimi S, Kushiki A, Guerguerian AM, Chau T (2012) Characterizing emotional response to music in the prefrontal cortex using near infrared spectroscopy. *Neurosci Lett* 525(1): 7–11.
76. Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, et al. (2010) Variation in transcription factor binding among humans. *Science* 328(5975): 232–235.
77. Schlattl A, Anders S, Waszak SM, Huber W, Korbel JO (2011) Relating CNVs to transcriptome data at fine resolution: assessment of the effect of variant size, type, and overlap with functional regions. *Genome Research* 21(12) 2004–2013.
78. Limb CJ, Braun AR (2008) Neural Substrates of Spontaneous Musical Performance: An fMRI Study of Jazz Improvisation. *PLoS ONE* 3(2): e1679.
79. Bengtsson S, Csikszentmihályi M, Ullén F (2007) Cortical Regions Involved in the Generation of Musical Structures during Improvisation in Pianists. *J Cogn Neurosci* 19:5: 830–842.
80. Perez I, Zatorre RJ (2005) Brain organization for music processing. *Annu Rev Psychol* 56: 89–114.
81. Blum K, Chen TJH, Chen ALH, Madigan M, Downs BW, et al. (2009) Do dopaminergic gene polymorphisms affect mesolimbic reward activation of music listening response? Therapeutic impact on Reward Deficiency syndrome (RDS). *Med.Hypotheses* 74: 513–520.